Molecular Informatics 2018

# Bimolecular Nucleophilic Substitution Reactions: Predictive Models for Rate Constants and Molecular Reaction Pairs Analysis

Polishchuk P., Antipin I., Varnek A.
*Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia*

## Abstract

© 2018 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim Here, we report the data visualization, analysis and modeling for a large set of 4830 SN2 reactions the rate constant of which (logk) was measured at different experimental conditions (solvent, temperature). The reactions were encoded by one single molecular graph – Condensed Graph of Reactions, which allowed us to use conventional chemoinformatics techniques developed for individual molecules. Thus, Matched Reaction Pairs approach was suggested and used for the analyses of substituents effects on the substrates and nucleophiles reactivity. The data were visualized with the help of the Generative Topographic Mapping approach. Consensus Support Vector Regression (SVR) model for the rate constant was prepared. Unbiased estimation of the model's performance was made in cross-validation on reactions measured on unique structural transformations. The model's performance in cross-validation (RMSE=0.61 logk units) and on the external test set (RMSE=0.80) is close to the noise in data. Performances of the local models obtained for selected subsets of reactions proceeding in particular solvents or with particular type of nucleophiles were similar to that of the model built on the entire set. Finally, four different definitions of model's applicability domains for reactions were examined.

http://dx.doi.org/10.1002/minf.201800104

## Keywords

bimolecular nucleophilic substitution reactions, Condensed Graph of Reaction, Generative Topographic Mapping, Matched Reaction Pairs, models applicability domain, Support Vector Regression

## References

[1] H. Kubinyi, QSAR: Hansch Analysis and Related Approaches, VCH, Weinheim, 2008
[2] A. Varnek, D. Fourches, F. Hoonakker, V. P. Solov'ev, J. Comput.-Aided Mol. Des. 2005, 19, 693–703
[3] T. R. Gimadiev, T. I. Madzhidov, G. Marcou, A. Varnek, Bionanoscience 2016, 6, 464–472
[4] T. I. Madzhidov, T. R. Gimadiev, D. A. Malakhova, R. I. Nugmanov, I. I. Baskin, I. S. Antipin, A. A. Varnek, J. Struct. Chem. 2017, 58, DOI 10.1134/S0022476617040023
[5] T. I. Madzhidov, A. V. Bodrov, T. R. Gimadiev, R. I. Nugmanov, I. S. Antipin, A. Varnek, J. Struct. Chem. 2015, 56, 1227–1234
[6] T. R. Gimadiev, T. I. Madzhidov, R. I. Nugmanov, I. I. Baskin, I. S. Antipin, A. Varnek, J. Comput.-Aided Mol. Des. 2018, 32, 401–414

[7] A. I. Lin, T. I. Madzhidov, O. Klimchuk, R. I. Nugmanov, I. S. Antipin, A. Varnek, J. Chem. Inf. Model. 2016, 56, 2140–2148

[8] A. G. Leach, H. D. Jones, D. A. Cosgrove, P. W. Kenny, L. Ruston, P. MacFaul, J. M. Wood, N. Colclough, B. Law, J. Med. Chem. 2006, 49, 6672–6682

[9] M. B. Smith, J. March, March's Advanced Organic Chemistry: Reactions, Mechanisms, and Structure, John Wiley & Sons, Hoboken, 2007

[10] A. A. Kravtsov, P. V Karpov, I. I. Baskin, V. A. Palyulin, N. S. Zefirov, Dokl. Chem. 2011, 440, 299–301

[11] F. Hoonakker, N. Lachiche, A. Varnek, A. Wagner, Int. J. Artif. Intell. Tools 2011, 20, 253–270

[12] T. I. Madzhidov, P. G. Polishchuk, R. I. Nugmanov, A. V Bodrov, A. I. Lin, I. I. Baskin, A. A. Varnek, I. S. Antipin, Russ. J. Org. Chem. 2014, 50, 459–463

[13] V. E. Kuz'min, A. G. Artemenko, E. N. Muratov, J. Comput. Aided. Mol. Des. 2008, 22, 403–421

[14] R. I. Nugmanov, T. I. Madzhidov, G. R. Khaliullina, I. I. Baskin, I. S. Antipin, A. A. Varnek, J. Struct. Chem. 2014, 55, 1026–1032

[15] Reilly JP, Hirata A (2016) Low-frequency electrical dosimetry: research agenda of the IEEE international committee on electromagnetic safety. Phys Med Biol 59

[16] ChemAxon., 2015

[17] ChemAxon., 2015

[18] T. I. Madzhidov, T. R. Gimadiev, D. A. Malakhova, I. R. Nugmanov, I. I. Baskin, I. S. Antipin, A. Varnek, J. Struct. Chem. 2017, DOI 10.15372/JSC20170402

[19] C. Muller, G. Marcou, D. Horvath, J. Aires-de-Sousa, A. Varnek, J. Chem. Inf. Model. 2012, 52, 3116–3122

[20] A. Dalby, J. G. Nourse, W. D. Hounshell, A. K. I. Gushurst, D. L. Grier, B. A. Leland, J. Laufer, J. Chem. Inf. Model. 1992, 32, 244–255

[21] EPAM Systems., 2015

[22] T. I. Madzhidov, R. I. Nugmanov, T. R. Gimadiev, A. I. Lin, I. S. Antipin, A. Varnek, Butlerov Commun. 2015, 44, 170–176

[23] J. Hussain, C. Rea, J. Chem. Inf. Model. 2010, 50, 339–348

[24] I. Sushko, S. Novotarskyi, R. Körner, A. K. Pandey, M. Rupp, W. Teetz, S. Brandmaier, A. Abdelaziz, V. V. Prokopenko, V. Y. Tanchuk, J. Comput.-Aided Mol. Des. 2011, 25, 533–554

[25] C. M. Bishop, M. Svensén, C. K. I. I. Williams, Neural Comput. 1998, 10, 215–234

[26] H. A. Gaspar, G. Marcou, D. Horvath, A. Arault, S. Lozano, P. Vayer, A. Varnek, J. Chem. Inf. Model. 2013, 53, 3318–3325

[27] D. Horvath, G. Marcou, A. Varnek, 2017, pp. 167–199

[28] J. Catalán, V. López, P. Pérez, R. Martin-Villamil, J.-G. Rodríguez, Liebigs Ann. 1995, 1995, 241–252

[29] J. Catalán, C. Díaz, Liebigs Ann. 1997, 1997, 1941–1949

[30] J. Catalán, C. Díaz, V. López, P. Pérez, J.-L. G. De Paz, J. G. Rodríguez, Liebigs Ann. 1996, 1996, 1785–1794

[31] R. W. Taft, M. J. Kamlet, J. Am. Chem. Soc. 1976, 98, 2886–2894

[32] M. J. Kamlet, R. W. Taft, J. Am. Chem. Soc. 1976, 98, 377–383

[33] M. J. Kamlet, J. L. Abboud, R. W. Taft, J. Am. Chem. Soc. 1977, 99, 6027–6038

[34] C.-C. Chang, C.-J. Lin, ACM Trans. Intell. Syst. Technol. 2011, 2, 27 : 1–27 : 27

[35] D. Horvath, J. Brown, G. Marcou, A. Varnek, Challenges 2014, 5, 450–472

[36] M. Mathea, W. Klingspohn, K. Baumann, Mol. Inf. 2016, 35, 160–180

[37] A. Varnek, D. Fourches, D. Horvath, O. Klimchuk, C. Gaudin, P. Vayer, V. Solov'ev, F. Hoonakker, I. V Tetko, G. Marcou, Curr. Comput.-Aided Drug Des. 2008, 4, 191–198

[38] I. V Tetko, I. Sushko, A. K. Pandey, H. Zhu, A. Tropsha, E. Papa, T. Oberg, R. Todeschini, D. Fourches, A. Varnek, J. Chem. Inf. Model. 2008, 48, 1733–46